

Understanding Machine Learning

So what is machine learning? This is actually a pretty tough question to answer if you've tried to Google it yourself. Because there's actually a lot of debate as to what M.L. is – like data science, statistics, AI, machine learning. They all cover similar topics, and people haven't agreed on what really machine learning should perfectly be defined as.

Some people think that machine learning is just automation, and that makes sense because a lot of times applications of ML are kind of like automation. For example, maybe you use a computer to transcribe an interview rather than a person, or you use something like Google translate to translate text that a person would have otherwise translated.

But actually for as long as programming has existed, programmers have been automating things. So, for example, maybe you have a floor-cleaning robot like a Roomba. Well, you don't actually need machine learning to make that thing work. You could just write some code that says that when the power buttons turned on, then have the thing sweep the floor. That is not machine learning.

Furthermore, machine learning can do more than just automate. For example, it can do things that humans can't do, like, for example, make a recommendation for what a reader might want to read based on what they've read in the past. Or forecast, I don't know, climate trends, or the price of bitcoin. Or make crazy artwork like this. How does that fit in with the pattern of what the rest of machine learning is?

Well, the way that I like to find ML is that it's just finding patterns in data, and usually using those patterns to make predictions about unseen data. You can think of it as sort of a new programming paradigm. So what does that mean? Well, let's take a look at classification. That's one of the types of tax people use machine learning for. One of the most common, so let's talk about classification.

Classification is just labeling things. So maybe you have pictures of dogs and cats, and you want to build an algorithm that given new pictures, can automatically label those dogs and cats for you. The way that we would do this traditionally with programming, before machine learning, is that a programmer would sit down and think what makes a cat a cat and a dog a dog, and they would come up with a bunch of different rules in their head. And then they would put this into their algorithms, so if an animal has a long tail and likes fish and doesn't like people, then it's a cat. Otherwise, it's a dog. That's the traditional programming approach.

What's different with machine learning is rather than a programmer coming up with rules, a machine learning algorithm learns by examples and comes up with rules itself. The way that this happens is that we build a machinery model. So first we have to collect lots of label training data, so hundreds or thousands of pictures of cats labeled cats and dogs labeled dogs. And hopefully many different breeds, and from different angles, and with different lighting. We show this labeled data to a machine learning algorithm, which then spits out a model.

So if your algorithm has done a good job, then when the model sees a new picture of a dog, it can label it as a dog. So this in a nutshell is what machine learning is. Or, at least that's what a classification algorithm is, which I told you is a big part of machine learning.

But again, it's sort of hard to really conceptualize all of the different ways that ML could be applied, so what I want to do is in the next few sections, is show you a bunch of different applications for machine learning used by newsrooms on all different sorts of data. And then hopefully you, like a machine-learning algorithm, can learn by example what's covered by this field.